Earth Sci Inform (2009) 2:63–74
DOI 10.1007/s12145-009-0026-7

RESEARCH ARTICLE

Virtual organisation in the SIMDAT meteorological activity: a decentralised access control mechanism for distributed data

Baudouin Raoult · Guillaume Aubert · Marta Gutiérrez · Cristina Arciniegas-Lopez · Ricardo Correa

Received: 12 September 2008 / Accepted: 15 April 2009 / Published online: 5 May 2009

Abstract The EU funded SIMDAT project is aimed at applying generic grid technology for the solution of complex application problems in several representative fields including automotive, aerospace, pharmaceutical and meteorology. To satisfy the requirements of the World Meteorological Organization (WMO) and the WMO Information Systems (WIS), the partners in the meteorology activity within SIMDAT (ECMWF, Deutscher Wetterdienst, the UK Met office, EUMETSAT and Météo-France), have developed grid-enabled software that provides generic distributed access to distributed meteorological data repositories via web-based portals, through a series of nodes organized in a mesh network. However, granting access to such an infrastructure, especially considering its fully distributed nature, is a serious challenge and a risk to the security of the overall grid infrastructure. SIMDAT solves this problem by implementing a security model based on a decentralized fine-grained access control mechanism that federates data providers and security issues using the notion of "trust domains". In this paper we highlight the main features of the SIMDAT grid application and describe its security model in detail.

Keywords Domain · ECMWF · Grid · ISO19115 · Mesh network · Metadata · Meteorology · Policies · Role · Security · VO

Communicated by H.A. Babaie

B. Raoult (✉) · G. Aubert · M. Gutiérrez · C. Arciniegas-Lopez · R. Correa
European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK
e-mail: Baudouin.Raoult@ecmwf.int

G. Aubert
e-mail: Guillaume.Aubert@ecmwf.int

M. Gutiérrez
e-mail: Marta.Gutierrez@ecmwf.int

C. Arciniegas-Lopez
e-mail: Cristina.Arciniegas-Lopez@ecmwf.int

R. Correa
e-mail: Ricardo.Correa@ecmwf.int

List of abbreviations

ACL	Access Control List
AMS	American Meteorology Society
AODV	Ad-hoc On-demand Distance Vector
CA	Certification Authority
DN	Distinguished Name
EDG	EU DataGrid
GSI	Grid Security Infrastructure
GT	Globus Toolkit
LAN	Local Area Network
IdP	Identity Provider
PKI	Public Key Infrastructure
RMI	Remote Method Invocation
RP	Resource Provider
SAML	Security Assertion Markup Language
SKOS	Simple Knowledge Organisation System
SP	Service Provider
SSL	Secure Socket Layer
SWEET	Semantic Web for Earth and Environmental Terminology
TLS	Transport Layer Security
VMC	Virtual Meteorological Centre
VO	Virtual Organisation
VOMS	Virtual Organization Membership Services
WIS	World Meteorological Organization Information System

WMO	World Meteorological Organization
WSDL	Web Services Description Language
WSS4j	Web Service Security Implementation for Java

Introduction

Data distribution and exchange are at the centre of core meteorological activities. Observation data is continually collected from distributed sources, fed into numerical weather prediction systems and re-distributed in raw format as well as in the form of prediction model outputs to forecast offices, service providers and end users. Collected observations of present validity are considered "real-time data" and are exchanged globally through dedicated networks to ensure that meteorological centres can feed these into forecasting models. Once the forecast is produced by the model, observations are stored into archives to ensure its preservation for post-analysis studies.

In the same manner, the forecast produced, of a present validity, represents real-time data usually available from online disks and it goes into archived mode once it passes its validity time. Data policies concerning a product change over its lifetime depending on whether this product represents its real-time status or the archived one. These policies might need to protect data of commercial value. Therefore, sharing of these datasets becomes a very delicate activity that needs to be carried out in a very controlled environment.

Other WMO designated centres and related programs, are also data producers wishing to publish and share the results of their studies and projects and ensure the long term preservation and accessibility to its data. In this vein, the forthcoming WMO Information System is a long term infrastructure initiative to ensure coordinated access and sharing of the data.

At present, there are well defined global policies recognised by the WMO data centres e.g. WMO Resolution 40, which relates to the provision of free and unrestricted exchange of data and products between WMO members and related programs. Although these data policies are well established and recognised by the different data centres, there is still a lack of infrastructure that allows sharing of data and policies in order to facilitate access to a user requesting data from multiple data centres. Research studies involving cross-boundary geo spatial locations result in multiple requests from a user who has to contact several meteorological centres and register at several institutes. Users would benefit considerably from a shared infrastructure that would allow them to retrieve data from any data centre without having to acknowledge many different policies and register at many different places in order to retrieve data from distributed locations.

Equally, data centres would benefit from an infrastructure that would allow them to publish and share datasets in a secured and well controlled environment that makes data available to end users and other data centres under dedicated agreements.

In order to support such activities, the meteorology activity of the SIMDAT project developed and deployed a virtual meteorological centre (VMC)—a common system with the aim of collecting and sharing distributed meteorological data (SIMDAT 2004b).

By using a grid infrastructure, the VMC provides the meteorological community with a single view of highly heterogeneous data kept in a variety of administrative and technological domains, systems and formats.

Virtual organisations

Resource federation leads to the virtualisation of all grid members into virtual organizations (VOs), where neither users nor resources are part of single organizations (Ahsant et al. 2006). VOs have been used as a bridge to establish *trust* relationships between the entity members (users, organizations, etc) and resource providers (RP) of a grid infrastructure. Thus, all entity members commit their resources to the VO and explicitly adhere to the common set of policies used to establish it (Foster et al. 2001). In return, VO members expect a trustworthy grid security model (Welch et al. 2003; Welch et al. 2005) that understands local security infrastructures, protects RPs assets and maintains user privacy but without limiting the VO's scalability and flexibility.

Consequently, research efforts within the grid community in recent years have focused on developing protocols, services and tools that help to build scalable VOs (Foster et al. 2001). In aiming to fulfil this vision, one of the most significant challenges for the grid computing community has been to develop a comprehensive group of mechanisms and policies for securing the grid (Welch et al. 2003; Humphrey et al. 2005; Ahsant et al. 2006). In short, any grid security model should be scalable, integrateable, interoperable, trusted and enforced. The need to support the integration and management of resources within the VOs has introduced complexity, challenging basic security issues (Welch et al. 2003).

In order to understand the different grid security architectures and technologies that have been developed, it is convenient to introduce a number of key concepts (Foster et al. 2001; Humphrey et al. 2005; IJISC 2006):

- Authentication: the process of verifying the identity of the user requesting access to a resource available at the VO.
- Credentials: the generic term to describe information provided by the user in order to be authenticated.

- X.509 certificates: electronic credentials used to identify a user, RP or organization that bind the identity's Distinguished Name (DN) with a public key.
- Authorization: the process of determining whether access to a resource should be granted to a certain user based on the policies written by the VO.
- Privacy: the enforcement of security policies to safeguard critical user information.

Authentication and authorization

Within a grid, the ability to perform remote work or requests is an essential task that requires the secure authentication and authorization of users (Snelling et al. 2004). During the early stages, grid security mechanisms were based only on the identities of the interacting entities (Foster et al. 2001; Sinnott et al. 2006b). A very common authentication method to establish a user identity within a grid is by using a Public Key Infrastructure (PKI), based on certificates issued by a trusted Certification Authority (CA). In the PKI model, entities are authenticated by presenting as credentials an X.509 digital certificate, which contains a unique Distinguished Name (DN) and a public key. The matching private key is kept securely by the entity. Authorization to access a specific resource is performed by using a locally managed Access Control List (ACL) of authorized users or *grid-mapfile*. As the *grid-mapfile* only allows the mapping users' DNs to local user names at resource level, it lacks the ability to grant fine grained access control to the resources, thus compromising the security of the providers (Sinnott et al. 2006a). Moreover, any access control relying only on the local listing names of authorized users is not scalable to the proportions needed for a VO that may grow dynamically (Sinnott et al. 2006a, b). Nevertheless, the PKI model has had widespread acceptance in the grid community.

Advanced security infrastructures

As grid applications move from scientific communities to more secured, focused areas, such as the health and industry sectors, this authentication-only security model is not only inapplicable but also hinders the VO scalability (Welch at al. 2003). Recently, considerable progress has been made in developing advanced security infrastructures that integrate authentication and authorization processes into the grid middleware (Foster et al. 2001; Demchenko et al. 2006; Sinnott et al. 2006b). Notable examples are the development of the Globus Toolkit[1] and the Unicore[2] middleware to enable grids making high performance computers available to scientific communities (Welch et al. 2003). In the Globus Toolkit, the Grid Security Infrastructure (GSI) module defines a unique credential set based on X.509 certificates issued by a trust CA, and a common protocol based on the transport layer security (TLS). The GSI credential set coupled with an associated private key is used by any grid entity to authenticate itself to other grid entities (Humphrey et al. 2005). As grids outside the research community can grow into thousands very quickly, this centralized model of certification based on CA goes against the VO principles of flexibility and scalability. Hence, the Globus Toolkit introduced the X.509 *proxy certificates* that allow a user (without CA participation) to assign dynamically a new X.509 identity to an entity and also delegate some subset of his/her rights to the receiver (Welch et al. 2003; Snelling et al. 2004). To manage a more fine-grained VO membership, the 3.2 Globus version introduced a VO-like Community Authentication Server (CAS) that issues certificates with access restrictions based on agreements between the VO and the resource owner (Pearlman et al. 2003). CAS coupled with proxy certificates is the attempt is the GSI's attempt to dynamically create and manage simple trust domains.

Another effort has been made by the EU DataGrid (EDG) test bed (Cornwall et al. 2003), in which the authorization process for each resource is managed by an automated procedure that derives local policies from central, manually managed source(s) of authorization (Alfieri et al. 2005). Neither the CAS nor EDG models are a scalable solution to potentially large grids. The EDG security model has been replaced by the Virtual Organisation Management Systems (VOMS), an authorization service that classifies VO users based on a set of attributes (Alfieri et al. 2004). VOMS differs from CAS and EDG, as it does not centralize all policy information (group membership and relevant rights) in a server (Alfieri et al. 2005). Instead, VOMS separates information about group memberships—stored in a server—from the information about granted rights stored and mapped at resource level (Alfieri et al. 2005). However, keeping a central VO user database places a large burden on each VO administrator as all grid users need to be added (Sinnott et al. 2006b). As a VO may expand rapidly and memberships change dynamically, the VOMS solution becomes impractical to administer (Welch et al. 2005).

The next generation of grids proposed the idea of *federation* for grouping users with similar properties or *roles*, in order to factor the *users* × *machines* problem into manageable parts (Chivers 2003). The meteorological activity can be seen as an example of federation, as it aims at grouping users wishing to access meteorological resources. As a federation is also able to group resources,

[1] Information on Globus Toolkit available at http://www.globus.org/toolkit/

[2] Information on Unicore Middleware available at http://www.unicore.eu/unicore/

it could be used as the foundation of a management system to control access to similarly federated resources (JISC 2006). Moreover, by using a federation the number of bipartite agreements between RPs and users can be minimized as all members of a federation sign up to an agreed set of policies (Chivers 2003) rather than having to establish bilateral agreements among all participant entities (Alfieri et al. 2004).

Fine-grained access control

Moving forward, the Shibboleth system proposed factoring grid security requirements into manageable domains using federation (Morgan et al. 2004). With Shibboleth, local organisations, known as Identity Providers (IdPs), are responsible for authenticating the user (by checking his/her credentials) and proving user's attribute information (JISC 2006). The decision to authorize access to a resource is made by the resource owner or Service Provider (SP), based on the user's attribute information (attributed-based authorization). Though deployed separately, the SP and IdP are used by Shibboleth to provide secure access to web-based resources (Morgan et al. 2004). Shibboleth is an implementation of the Security Assertion Mark-Up Language (SAML), which provides interoperability among web sign-on products. Furthermore, the GridShib project (Welch et al. 2005) is an attempt to provide a mechanism whereby a grid service authenticates a user using the Globus GSI, and authorization is performed by using user attributes from the Shibboleth IdP.

With the emergence of grid services, some research has focused on grid scalability issues, and on implementing federated security services and policy-based access control systems (Ahsant et al. 2006; Demchenko et al. 2006; Lang et al. 2006; Sinnott et al. 2006b). Advance authorization services aiming at fine grained access control are based on users sharing similar roles (Demchenko et al. 2006), for instance membership of the same project, all researchers at ECMWF, etc. As Sinnott et al. 2006a state attribute based-authorization is scalable to global proportions and when it is coupled to a policy based authorization system, allows fine-grained access control to local resources. One of the challenges of the SIMDAT architecture is, therefore, to embrace multi-organization federations, allowing scalable growth of the distributed infrastructure. Additionally, the SIMDAT's VMC security architecture needs control mechanisms to protect users' privacy, in order to comply with local legal security requirements.

VO in the SIMDAT meteorological activity

In this paper we outline a novel framework that organizes secure access to distributed users and datasets that are part of a dynamic VO. In doing so, SIMDAT introduces the concept of *domain,* in order to federate resources from the participant members of the VO. The proposed security framework implements an attribute based access control model, where authentication and authorization are two separate processes that depend on the *trust relations* established in each *trust domain*. The framework is explored in the context of a distributed, scalable, flexible and loosely coupled grid architecture that expects the provision of a unified but highly secured view of meteorological data. The proposed approach and solution have been developed to respond both to general security issues regarding grid infrastructures and to the specific requirements of the meteorological community. Nevertheless, many other communities aiming to share resources using a grid data structure may benefit not only from the experience but also from the technological solution developed.

The Virtual Meteorological Centre (VMC)

The VMC is an innovative, decentralised framework deploying a grid infrastructure that offers cataloguing, discovery and data retrieval facilities to the meteorological community. Moreover, to overcome the multiple security challenges in any grid environment (Chivers 2003; Welch et al. 2003; Humphrey et al. 2005; Sinnott et al. 2006a) VMC incorporates a novel security model that protects users' privacy and data resources, but without hindering the basic grid principles. The VMC platform is the result of a joint effort by the meteorological centres of Germany (Deutscher Wetterdienst), France (Météo-France) and the United Kingdom (UK MetOffice) and two European organisations: ECMWF (European Centre for Medium-range Weather Forecasts) and EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites).

A three tier component stack

VMC integrates a stack of three self-contained and loosely coupled components: the portal, the catalogue node (CN) and the data repository (DR). The VMC architecture, as shown in Fig. 1, is assembled by interconnecting a series of nodes through a dedicated, secure communication layer. Through a web-based portal (pink component in Fig. 1), each node is able to provide distributed access to data held in legacy systems bridged by the DR interface (green panel in Fig. 1).

The VMC solution offers a decentralized virtual catalogue made up of standard metadata records describing the resources held by each of the meteorological centres or RPs. The heterogeneous resources owned by diverse data publishing centres are thus federated into a single view or

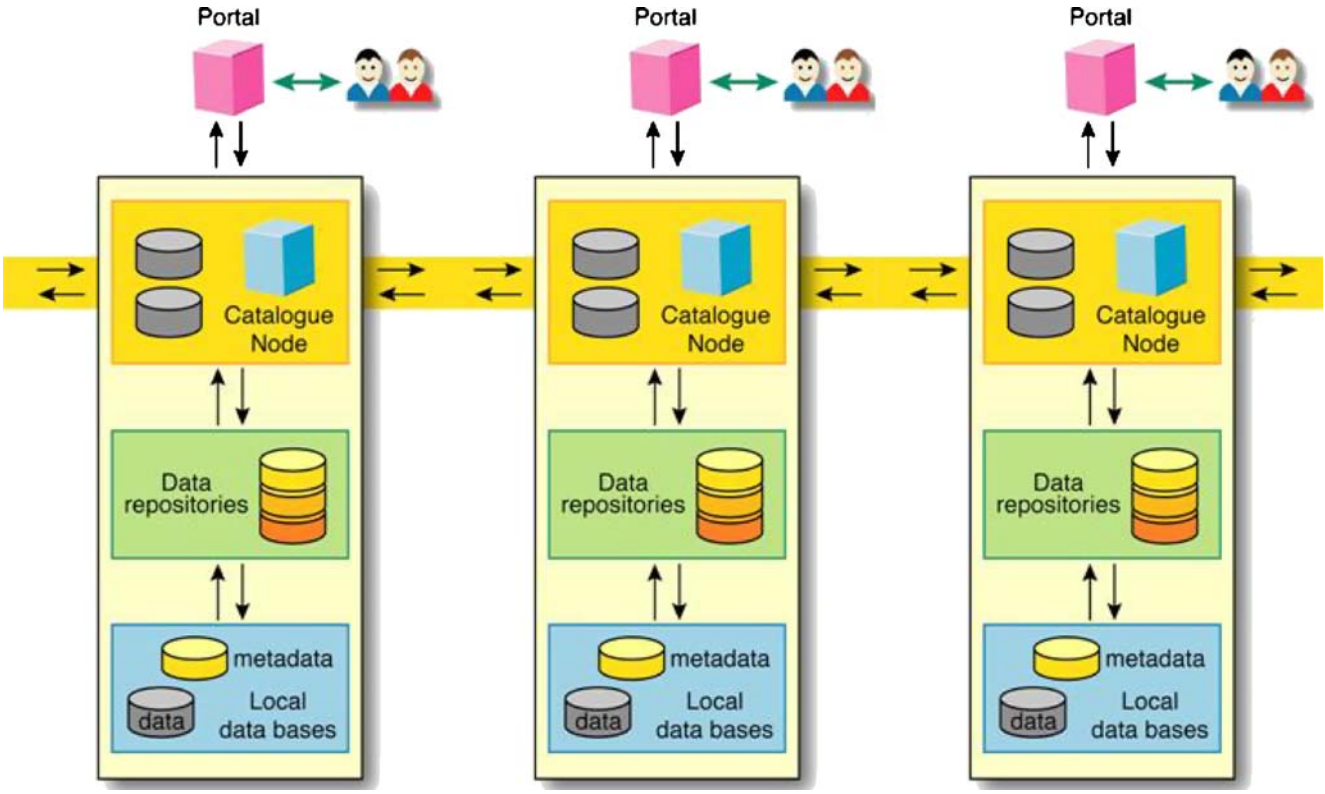

Fig. 1 Three tier VMC architecture

virtual catalogue made available via the VMC infrastructure. Consequently, the VMC architecture provides the right support for building a distributed network not only of meteorological centres but of any group of providers interested in offering data and resources to a wider community.

As the Fig. 1 illustrates, the portal provides the VMC with a flexible and user-friendly, web-based interface to discover, request, subscribe to or download data resources from the virtual catalogue. The portal's implementation exploits metadata- and ontology-based technologies to provide users with a more intelligent searching and browsing service over the virtual catalogue. The discovery services incorporates a scoring system that creates indexes from relevant metadata fields, such as title, abstract, keywords, temporal and geographical coordinates, etc. In addition, by incorporating earth science ontologies such as GMET and SWEET (Pouchard et al. 2003), the portal provides users with a controlled navigation tool, based on the taxonomical relations between meteorological terms, thus increasing the probability of a more successful, rapid and efficient data discovery.

To offer a comprehensive discovery service at the portal, the VMC architecture required a component that maintained the virtual catalogue. Consequently, the CN was designed to provide connectivity amongst RPs and to communicate with local data sources via the DR(s). The VMC solution implemented a synchronization engine, run by the CNs, that ensured every node had a copy of the global metadata catalogue, and provided users with fast responses when searching for datasets.

A noteworthy feature of VMC is that it offers a non-instructive infrastructure that is easily adaptable to the existing legacy systems held by different data providers (meteorological services). To achieve this, the VMC has been provided with the DR, a unified interface that bridges the CN to any existing data distribution system. The DR provides the facilities to translate on demand data subscription requests into requests that the interfaced data source understands. The DR also acts as a metadata provider by describing any accessible dataset according to the ISO19115 standard.

The infrastructure architecture

One of the main limitations of an early VMC prototype was the fully connected nature of the grid architecture. This implied that each node had to be connected to every other node, in order to get a full copy of the catalogue or retrieve datasets from any site. As the number of participating nodes gradually grew, the fully connected architecture proved neither scalable nor manageable. Then, the VMC was improved by implementing a fully decentralized architecture, following a peer-to-peer model (Iamnitchi et al. 2002).

The solution jointly developed with INTEL, consisted of adopting algorithms and technology used in the mobile telephony world to build a mesh network. Within a mesh network, each peer is connected to a small number of peers and dedicated routing algorithms were incorporated to determine the fastest route to a particular peer destination. Two implementations of the routing service have been integrated: a static map routing which, based on a static map available at each peer, calculates "peer-to-peer routes" following the Dijkstra Shortest Path Algorithm (Dijkstra 1959); and a more dynamic approach, implemented by using the Ad-hoc On-demand Distance Vector—AODV Algorithm (Perkins and Royer 1999), which establishes a route to a destination only on demand, hence reducing the network traffic. The router service has the capability to forward requests using a chain of intermediate peers within the mesh network. As the VMC architecture is fully decentralised, there is not a central point of failure. Moreover, partners can be added to and removed from the grid, without having a major impact on the network connectivity.

Securing the VMC infrastructure

From an administrative viewpoint, the VMC was established as a set of organizations, people and resources with the common goal of offering integrated meteorological data and value-added services (SIMDAT 2004a). Thus, the partners in the SIMDAT meteorological activity agreed on establishing a VO within the context of the meteorological domain, to develop and implement an operational framework to manage the VMC and to establish mechanisms and policies for securing the grid.

Security challenges within the VMC infrastructure

As the mesh network did indeed prove to be a scalable and flexible decentralized architectural solution for the VMC, the authorization and authentication process within the security model needed to match the peer-to-peer environment, without introducing any single point of failure.

Within the VO context, each VMC partner maintains its own user registration and implements its own policies for accessing resources. Thus, as the DR provides a non-intrusive way to bridge any legacy systems, so the security model needed to deliver an authorization mechanism that allowed each resource provider (DRs owners) to retain ultimate control over the policies that govern access to its resources (Pearlman et al. 2003). The security solution therefore requires that users are authenticated at one site (the portal) and authorised at another (the DR), without exchanging user information. This requirement for user privacy is dictated by national laws such as the "*Data Protection Act 1998*"[3] in the UK and by the "*Loi informatique et libertés*"[4] in France.

The VMC-VO trust model

Trust management, as Humphrey et al. 2005 defined, is the process of deciding what entities are to be *trusted* to do what *actions*. Hence, the VMC-VO adopted the concept of *trust* relationships amongst the participating entities as the building block of its security architecture. All the entities involved in such an agreement formed a *domain of trust* or *domain* for short.

A *domain* is defined as a group of organizations that share a common set of roles and data access policies. All members of the domain must have the same understanding of each of the roles and data policies that they share. This is achieved by "out-of-bound" agreements, that can take the form of a contract (e.g. between members of a project), a convention (e.g. between ECMWF and its members) or a resolution (e.g. amongst all Member States of the WMO), etc. Once this has been established, all members of a given *domain* trust each other for allocating roles and authorizing users to access any data that they make available on the grid.

Within the VMC-VO, an organisation can be part of several *domains*, thus sharing different datasets with different partners, while still being part of the same infrastructure. An example of such domains could be all national meteorological services represented at the WMO, which all agree on a common data policy defining what data are freely available for research and education. Having established a common understanding of the definition of "research", one meteorological service would deliver the agreed data to a user, if another meteorological service had claimed that this user was a researcher.

Implementation

Within the VMC, a *domain* delimits the scope of the data polices of the organizations that belong to the VO. Consequently, a *domain* defines a group of sites with a common data access policy. The domain's member entities have to define the data access policies visible to that domain only. These *policies* are expressed in terms of *roles*. Roles and data policies are interchangeable terms, e.g. the role researcher and the data policy researcher (see Fig. 2).

[3] https://www.uktradeinvest.gov.uk/ukti/appmanager/ukti/help?_nfls=false&_nfpb=true&_pageLabel=data_protection_act
[4] http://www.cnil.fr/index.php?id=4

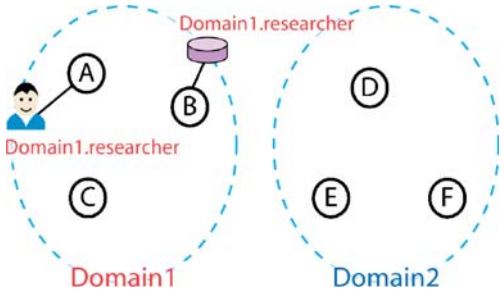

Fig. 2 Mapping data policies with users' roles

In order to establish a domain, all members exchange their public keys, which will later be used to check the authenticity of the sites' signature attached to any request.

Within a domain, users may have several roles and a dataset may have several associated data policies. Thus to access a given dataset, a user must have at least one *role* that matches at least one data *policy* of that dataset. In the figure above, B publishes a dataset with the domain1. researcher data *policy* and a user at A has been assigned to the domain1. researcher *role*. As the role and data policy match and A and B belong to the same domain domain1, the user at A is granted access to the relevant dataset at B.

By mapping *role* × *data policies*, the VMC-VO security architecture was able to factor the *users* × *machine* problem (Chivers 2003) into manageable parts. Instead of having to establish $n \times n$ number[5] of bi-lateral agreements amongst all the entities within the *domain,* a newly joining organization has only to agree on the existing *role* × *data policies* for that specific *domain*. This decision resulted from the realisation that, in any distributed system involving several organisations, there must be a formal agreement that defines a common understanding of the roles/data policies.

Implementing the trust domain-based security model posed a series of design and technological challenges to the SIMDAT developers. The VMC implementation supports multiple domain membership, thus a site can belong to many other domains. However, in order to successfully manage the complexity and enable grid scalability and growth, domains in the VO do not form a hierarchy but a flat structure. The VMC avoids policy or role name clashes amongst domains by implementing domain names as "name spaces" for roles and data policies. Thus, the role research in the domain1 domain differs from the role research in the domain2.

Prior to establishing a domain, potential members must agree on a list of roles/data policies for accessing the resources federated by the domain. Domain authentication is therefore managed by exchanging public keys between the domain members, so each site holds a repository that associates the domain with a list of public keys. Adding a key to this repository is the responsibility of the site administrator.

In order to publish data, the VMC allows that the corresponding data policies (including one or more domain name spaces) are attached to the dataset metadata. In this way, users know, while browsing the catalogue at the portal, whether their current credentials grant access to a particular dataset. Moreover, one dataset can be published with different data policies in different domains.

Putting it all together

Figure 3 provides an overview of the way the security model has been implemented within the VMC three tier architecture.

The *Login Service* allows a user to log-in at any home site using the portal web-based interface. As illustrated in Fig. 3, the user is authenticated at the first point of contact.

When a user requests a specific dataset from a remote site (right hand side node), the user home site (left hand side node) issues a request. The user roles are attached to the request that is signed by the user home site using its private key. The request is forwarded to the destination node by a chain of intermediate nodes according to the topology of the mesh network. Intermediate nodes may not need to be in the same domain as the sender node to forward the request.

The destination site receives the request and checks its signature (sender site's private key) against the list of known domain members using the public keys previously stored. If the request is signed by a known site, the roles from all the domains that are common to sender and receiver sites are extracted from the request. This list is matched against the data policies of the requested dataset. If any of the user's roles matches any of the dataset policies, access to the data is granted.

Scenario 1: Authentication and authorisation within a domain

The following scenario illustrates a user login sequence, using the access control model based on trust domains proposed by the VMC (Fig. 4).

1. The B site publishes a dataset with the domain1. researcher data policy.
2. A user is registered with the domain1.researcher user role established within the domain1.
3. The user logs into A and is authenticated by proving his/her identity (X.509 certificate, password and username). The user wants to access a dataset in B.

[5] Where "n" is the total number of entities within the domain.

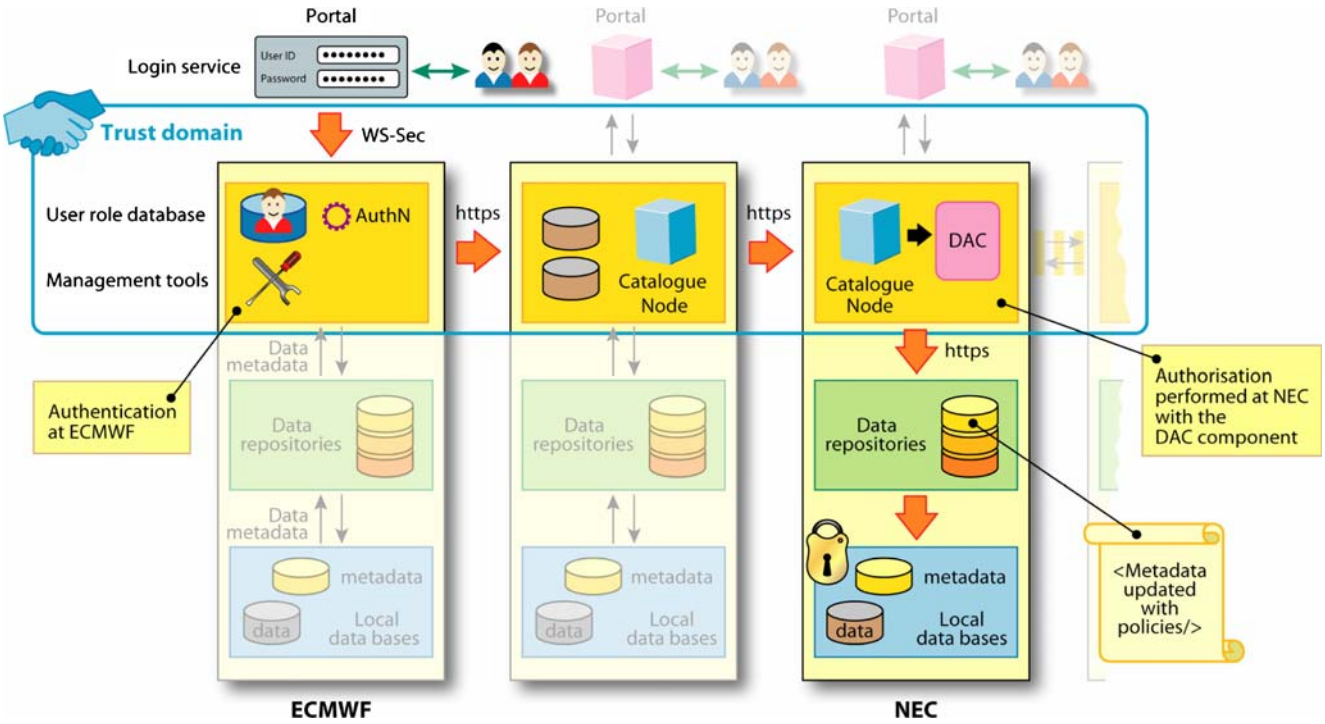

Fig. 3 The proposed security model and the VMC architecture

4. A assembles a request using the domain1.researcher user role and signs it using its private key. Then, A sends the assembled request to B.
5. B checks the A signature against the known public keys. It also checks if A is member of the domain1. As both belong to the same domain, B trusts A. Finally, B checks the *role* within the received request against the data *policy*. If they are the same, B grants the user access to the requested dataset.

Scenario 2: Handling requests from other domains

This scenario describes how the requests coming from other domains are handled by the VMC (Fig. 5).

1. The B site publishes a dataset with the domain1. researcher data policy.
2. A user is registered with the domain2.researcher user role established within the domain2.
3. The user logs into D and is authenticated. The user wants to access a dataset in B that belongs to the domain1.
4. D assembles a request using the domain2.researcher user role and signs it using its private key. Then, D sends the assembled request to B.
5. B checks the D signature against the known public keys. As D is not a member of domain1, access to the data is denied, without even considering the roles.

This scenario also demonstrates that, if a remote site were to pretend that one of its users had a valid role from the domain1, access to the data would still be denied, as the request would not be signed by a trusted member of the domain.

Scenario 3: Access through a different site

The third scenario shows that the proposed security model will allow users to have access to the grid, even if

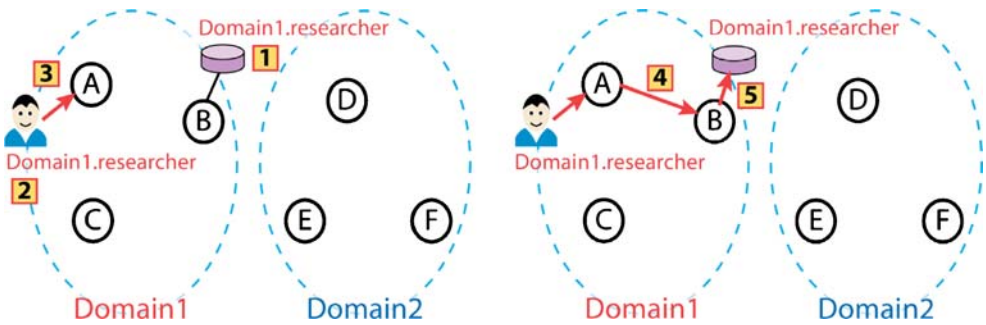

Fig. 4 Authorization and authentication within the same domain

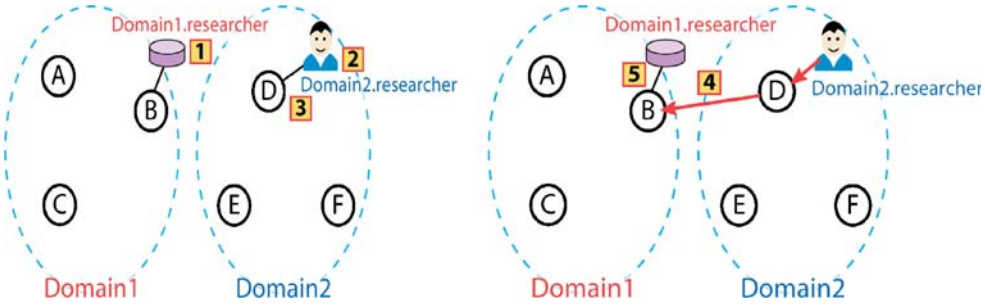

Fig. 5 Handling requests from other domains

their home site is not available. This can be achieved without the need for users to register at several sites or to exchange any user specific information between the nodes (Figs. 6 and 7).

1. The B site publishes a dataset with the domain1. researcher data policy.
2. The user logs into A and is authenticated by proving his/her identity (X.509 certificate, password and username). The user wants to access a dataset in B.
3. A exports the user's credentials (e.g. the user's identity and a list of roles) into a "wallet" and signs it using its private key.
4. A suddenly becomes unavailable.
5. The user is able to login at C by using his "wallet".
6. C accepts the user's login because it has been signed by A which belongs to the same domain. The user's authorization is performed by checking that the wallet has been signed by a node from the same trust domain.
7. Now, the user can issue a data request to B. Before sending it, C signs it with its private key and also attaches the roles signed by A from the original request.
8. Finally, B checks that both A and C signatures come from the same trusted domain. This is performed before B checks the user's role against the dataset data policy.

Results and further work

The VMC infrastructure has been installed and tested by nine different meteorological centres all over the world (see Fig. 8); it successfully provides 24/7 access to a virtual catalogue comprising more than 27,000 datasets distributed amongst all its partners.

The architecture has proved to be very flexible, as it allows different deployments, which can adapt to the different network topologies and security constraints of each of the partners. The adapted mesh network topology provides a very scalable infrastructure to the VMC, as a CN needs only to be directly connected to a limited number of peers.

On top of this architecture, security modules following the model described were implemented following standard GRID and SOA technologies.

The authentication service is based on Web Services technologies, making use of WS-Security between the portal and CN interfaces.

The authorisation and trust services are built-in modules within an XML based message SOA framework that communicates the CN peers. Data requests carry Security Assertion Markup Language (SAML) structures asserting authenticated users' requests, their roles within a domain and domain membership for the issuing organisation. SAML assertions allow users' requests to reach a remote node and request data without the need for authenticating again, at the remote site. The trust established between data centres allows for authenticated users to place requests within the domain of trust.

PKI infrastructure and W3C XML Signature are used to digitally sign the requests that travel to remote peers where verification modules check for the integrity of the message as well as the authenticity of the issuer. PKI is also used to establish secured communication between the CN channels to send messages encrypted.

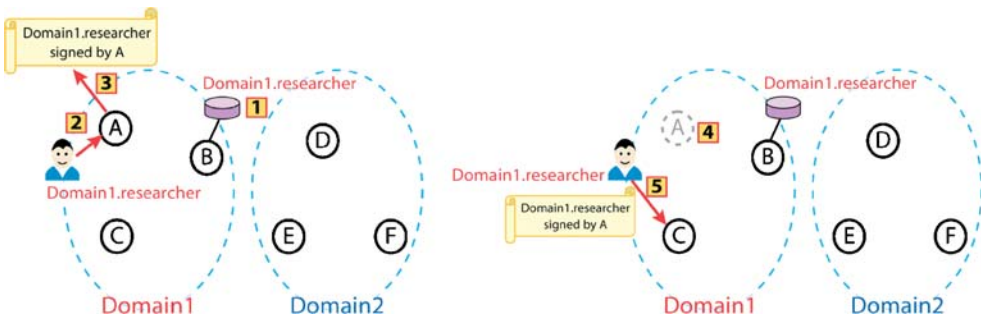

Fig. 6 Redirecting authorization when the user's home site is unavailable (part one)

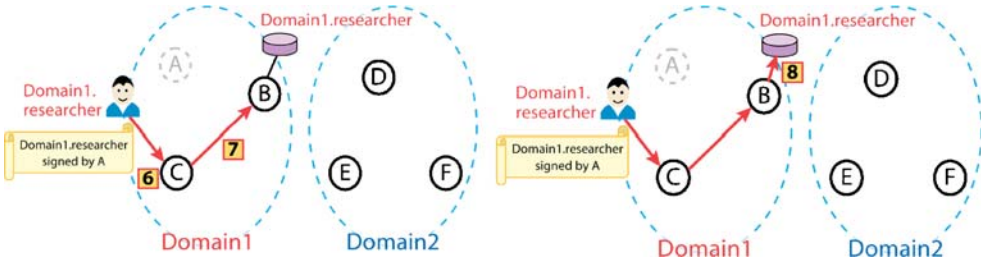

Fig. 7 Redirecting authorization when the user's home site is unavailable (part two)

Data policies are expressed technically in terms of roles and these are uniquely identified within a domain. The domain qualified roles are described as part of the metadata describing a dataset. These are checked against the user's roles when a data request and its SAML assertion containing the role information arrives from a remote site.

The software has achieved its maturity implementing distributed access controlled to protected datasets in trusted domains. On the other hand, the implementation is still lacking from scenario number 3, described in the previous section, in order to support end users wanting to use any portal from a trusted domain as a back up mechanism.

Finally, the establishment of trust and VO formation is a manual process. The organisations interested in establishing a domain of trust, negotiate "out-of-band" the policies and terms for accessing the VO as well as the individual membership. Policies, domains and trusted organisations are then catalogued and imported at each of the individual members, of course this information is not synchronised and stays at each of the data centres' own catalogues. One of the drawbacks that arises from this manual set up comes when domain or policy information needs to be updated. Under these circumstances, the framework would benefit from a notification mechanism and automatic ingestion of new or updated policies at the trusted sites.

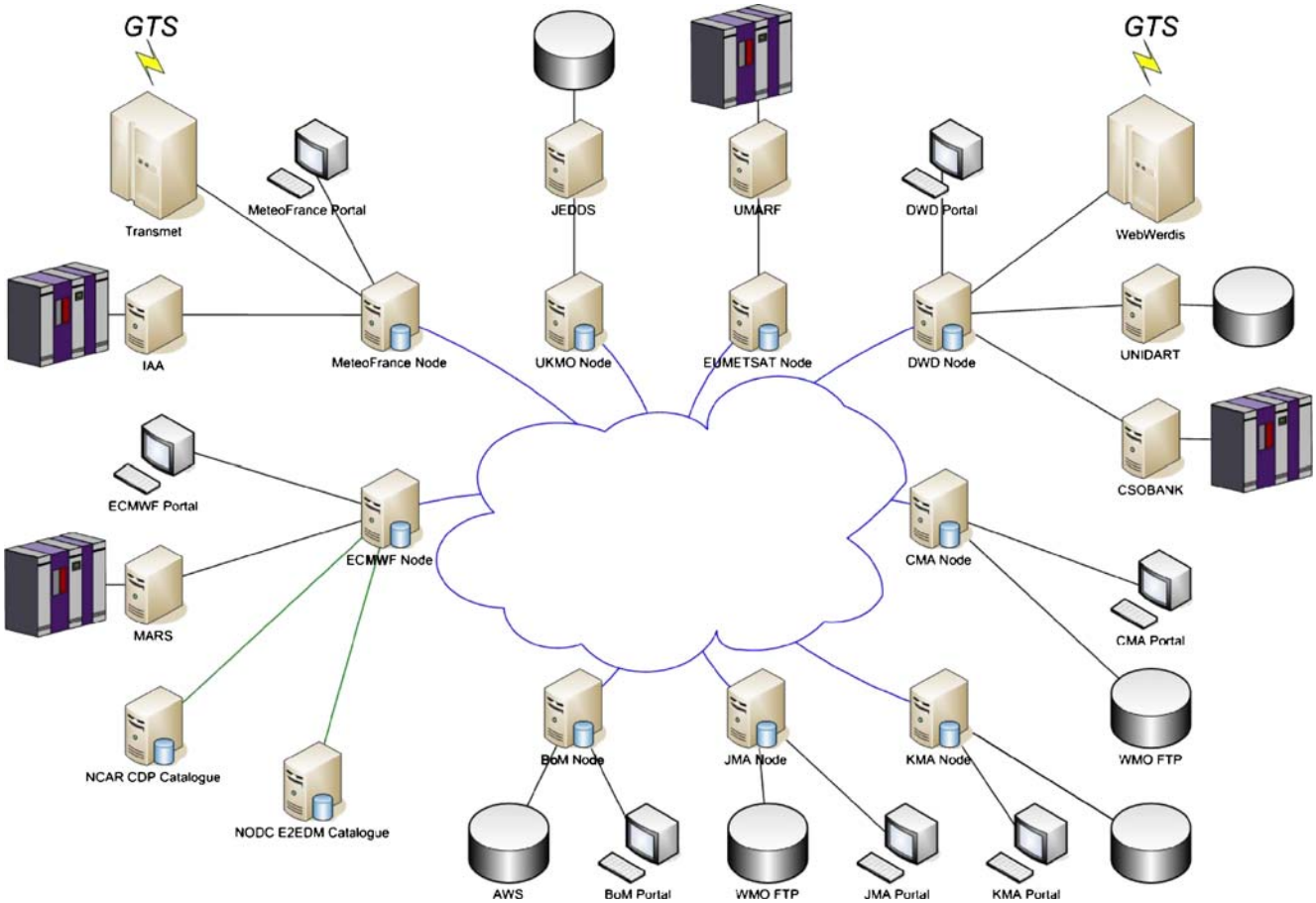

Fig. 8 VMC data grid

Discussions and conclusions

Studies of existing security models outlined the limitations of available grid middleware in handling fine-grained access rights to distributed data in a way that protected not only available resources, but also users' privacy, without hindering local sites' autonomy (Chivers 2003; Alfieri et al. 2005; Humphrey et al. 2005).

The security model proposed in this paper not only allows a fine-grained access control mechanism to the available resources, but has also proved to be a scalable solution that fulfils the requirements of the 24/7 distributed robust architecture needed for sharing meteorological data. The proposed security model has provided the VMC architecture with authentication and authorization components that can be easily interfaced within any existing security systems at the organization or data providers involved.

The VMC provides data centres with a framework that allows them to publish and share data in a secure environment with minimal disruption to their existing legacy systems. The establishment of trust domains allows for the creation of well defined policies that pertain to data defined only within the domain. Thus, adoption of the framework either to enforce global policies e.g. such as WMO Resolution 40, or to enforce policies at a smaller scale domain would allow for automatic sharing of secured and distributed datasets.

At the same time, end users can make use of a portal facility that provides them with a single view catalogue of data distributed in multiple data centres. Users can then retrieve datasets from multiple locations without the need for registering at every data centre from which they require data. The VMC allows users distributed data retrieval in a seamless and secure way without exposing the end user to the complexities of a distributed platform.

Acknowledgments We would like to acknowledge the help and support of all the SIMDAT partners, in particular the personnel of DWD, the UK Met Office, Météo France and EUMETSAT who have been directly or indirectly involved in this project. We would also like to thank the Bureau of Meteorology (Australia), the Chinese Meteorological Agency (CMA), the Japanese Meteorological Agency (JMA), the Korean Meteorological Agency (KMA), the National Center for Atmospheric Research (NCAR, USA) and the Russian National Oceanographic Data Centre (RNOCD) for their collaboration with the project.

References

Ahsant M, Surride M, Leonard T, Krishna A, Olle M (2006) Dynamic trust federation in grids. Lecture Note in Computer Science 3986:3–18

Alfieri R, Cecchini R, Ciaschini V, dell' Agnello L, Frohner Á, Gianoli A, Lörentey K, Spataro F (2004) An authorization system for virtual organizations. In Proceedings of the 1st European Across Grids Conference. Santiago de Compostela, Spain, pp 33–40

Alfieri R, Cecchini R, Ciaschini V, dell' Agnello L, Frohner Á, Lörentey K, Spataro F (2005) From Gridmap-File to VOMS: managing authorization in a grid environment. Future Gener Comput Syst 21(4):549–558

Chivers H (2003) Grid security: problems and potential solutions. In UK e-science programme, March 2003. http://www.cs.york.ac.uk/ftpdir/reports/YCS-2003-354.pdf. Accessed 6th Aug 2008

Cornwall L, Jensen J, Kelsey D, McNab A (2003) EU datagrid and gridPP authorization and access control. In Proceedings of the UK e-Science All Hands Meeting 2003, pp 382–384, EPSRC

Demchenko Y, Gommans L, Tokmakoff A, van Buuren R (2006) Policy based access control in dynamic grid-based collaborative environment. Proceedings of the International Symposium on Collaborative Technologies and Systems (CTS), pp 64–73

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1:269–271

Foster I, Kesselman C, Tuecke S (2001) The anatomy of the grid: enabling scalable virtual organizations. Int J Supercomput Appl 15(3):200–222

Humphrey M, Thompson MR, Jackson KR (2005) Security for grids. Proc IEEE 93(3):644–652

Iamnitchi A, Foster I, Nurmi DC (2002) A peer-to-peer approach to resource location in grid environments. In Proceedings of the 11th Symposium on High Performance Distributed Computing, Edinburgh, UK

JISC (2006) Shibboleth: connecting people to resources. http://www.jisc.ac.uk/uploaded_documents/JISC-BP-Shibboleth.pdf. Accessed 6th Aug 2008

Lang, B, Foster I, Siebenlist F, Ananthakrishnan R, Freeman T (2006) A multipolicy authorization framework for grid security. Proceedings of the Fifth IEEE International Symposium on Network Computing and Applications, pp 269–272

Morgan RL, Cantor S, Carmody S, Hoehn W, Klingenstein K (2004) Federated security: the shibboleth approach. EDUCASE Quarterly 27(4):12–17

Pearlman L, Kesselman C, Welch V, Foster I, Tuecke S (2003) The community authorization service: status and future. In Proceedings of International Conference for Computing in High Energy and Nuclear Physics

Perkins CE, Royer EM (1999) Ad hoc on-demand distance vector routing. In Proceedings of the 2nd IEEE workshop on mobile computing systems and applications. New Orleans, LA, USA, pp 90–100

Pouchard L, Cinquini L, Drach B, Middleton D, Bernholdt D, Chanchio K, Foster I, Nefedova V, Brown D, Fox P, Garcia J, Strand G, Williams D, Chervenak A, Kesselman C, Shoshani A, Sim A (2003) An ontology for scientific information in a grid environment: the earth system Grid. In: Proceedings 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid CCG 2003, pp 626–632

SIMDAT (2004a) D 19.1.2 Report on consolidated technology prototypes for Virtual Organisations and Ontologies. http://www.ecmwf.int/services/grid/simdat/wiki/do/get/documents Accessed 8th September 2008

SIMDAT (2004b) D.18.1.1 Consolidated Meteorology Requirements. In Grid based Systems for solving complex problems—SIMDAT Project. http://www.scai.fraunhofer.de/fileadmin/images/nuso/SIMDAT/SIMDAT_D.18.1.1_public.pdf. Accessed 4th Aug 2008

Sinnott RO, Chadwick DW, Koetsier J, Otenko O, Watt JP, Nguyen TA (2006a) Supporting decentralized, security focused dynamic virtual organizations across the grid. In Proceedings of the Second IEEE International Conference on e-Science and Grid Computing (e-Science'06)

Sinnott RO, Ajayi O, Stell AJ, Watt J, Jiang J, Koetsier J (2006b) Single sing-on and authorization for dynamic virtual organizations. In: Camarinha-Matos LM, Afsarmanesh H, Ollus M (eds)

Network-Centric collaboration and supporting frameworks, vol 224. Springer, Boston, pp 555–564

Snelling D, van den Berghe S, Qian Li V (2004) Explicit trust delegation: security for dynamic grids. FUJITSU Sci Tech J 40(2):282–294

Welch V, Siebenlist F, Foster I, Bresnahan J, Czajkowski K, Gawor J, Kesselman C, Meder S, Pearlman Laura, Tuecke S (2003) Security for grid services. High performance distributed computing, 2003. In Proceedings. 12th IEEE International Symposium, pp 48–57

Welch V, Barton T, Keahey K, Siebenlist F (2005) Attributes, anonymity, and access: shibboleth and globus integration to facilitate grid collaboration. In Proceedings of the 4th Annual PKI R&D Workshop: Multiple paths to trust, April 2005, pp 15–15